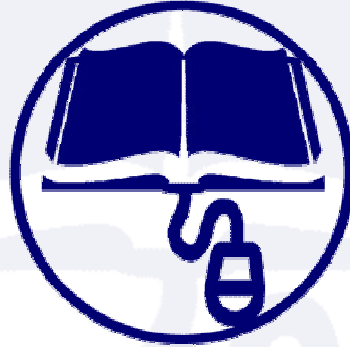

Herramientas y Estándares para las Bibliotecas Digitales del Nuevo Milenio



Rodrigo Torrén Heeren

Unidad de Teleinformación

Universidad de Los Andes

Mérida - Venezuela

torrens@cecalc.ula.ve

Puebla, México, Mayo 2002

Introducción

- Existe mucho interés hoy en día en las llamadas “Bibliotecas Digitales” ...

(por eso estamos aquí hoy por ejemplo)

- Una búsqueda en “Altavista” con las palabras...

“Digital Libraries”

...produce mas de 120.000 respuestas...

- Existen cientos de instituciones estudiando y generando proyectos e información relacionada con el tema....

¿Por qué? ... ¿Para qué Bibliotecas Digitales?

...

Objetivos de Desarrollar Bibliotecas Digitales

Hay una necesidad grande de acceder y estudiar colecciones de información provenientes de instituciones geográficamente dispersas.

Se deben hacer esfuerzos de integración a la...

"Infraestructura de Información Global"

(Global Information Infrastructure)

(ISTEC-LibLink, Digital Libraries Workshop. Albuquerque NM 1999)

...Objetivos de Desarrollar Bibliotecas Digitales

Según la iniciativa Library Linkages (LibLink) del ISTECS se pretende:

- Fomentar la distribución de información a través de conexiones de redes.
- Desarrollar bases de datos regionales.
- Mejorar la disponibilidad de material para la educación e investigación.

Necesidades

Grupos productivos necesitan comunicarse e intercambiar información, estableciendo lo que se ha llamado:

Ambientes de Investigación Digitales
(*"Digital Research Environments"*)

...Necesidades

Esfuerzos a varios niveles:

- Individuales
- Grupales
- Organizacionales
- Nacionales
- Regionales
- Globales

...Necesidades

Hacen falta:

- Idiomas estándares de representación y almacenamiento de datos
- Protocolos de comunicación estándares
- Protocolos estándares de consulta y recuperación de información
- Estándares de representación de conocimientos (Knowledge Representation)

Principios para construir Bibliotecas Digitales

- Espere el cambio
- Conozca sus contenidos
- Involucre a la gente correcta
- Diseñe sistemas usables
- Asegure el acceso libre-abierto
- Tenga cuidado con los derechos de los datos
- Automatice donde sea posible
- Adopte y adhiera a estándares
- Asegure la calidad
- Tenga cuidado con la persistencia de los datos

(Comm. Of the ACM. Mayo 2001)

Nos concentraremos aquí en:

Lenguajes y Estándares como Herramientas

Lenguajes de Marcas

Los “lenguajes de marcas” (*Markup Languages*) como HTML, SGML y XML, usan “etiquetas” u otros indicadores para decirle a la computadora como mostrar, reconocer, o manipular de cualquier manera, material “marcado”.

```
<?XML version="1.0" rmd="all"
encoding="utf8" ?> <!DOCTYPE tei.2
PUBLIC "-//TEI//DTD Historical
Editions//EN"> ...

<p>to recapitulate ... the greivances
which America <add>has long</add>
labour'd under & which after
<del>long</del> <add>unexampled</add>
patience & forbearance on her part
...

...

<h1><a name="ID8">Why isn't SGML
universal?</a></h1><p>
SGML solves the problems of the Web,
but ...

<UL>
<li>undeniable reputation for
complexity</li>
<li>parsers tend to be large and hard
to produce</li>
<li>hence they tend not to be
incorporated into other pieces of
software like Web browsers</li>
</UL>
```

HTML

- El *Hypertext Markup Language* (HTML) es el lenguaje de marcas más difundido.
- Casi todos hemos tenido algún contacto con él
 - *Recuperando información que usa este formato*
 - *Produciendo documentos HTML*
- Lenguaje con estructura muy simple (ventaja), orientado a

PRESENTACIÓN

- Funcionalidad muy limitada
- Diferentes versiones y soporte por parte de navegadores

...HTML

- La limitación real de HTML es su incapacidad para...

...***describir el contenido de un documento***

- HTML fue construido para “publicación” no para facilitar la recuperación o intercambio de información.
- Debido a su simplicidad, seguirá siendo el lenguaje escogido por muchos para codificar texto y gráficos así como para composición y formato de presentación de documentos.

XML

eXtensible Markup Language

- La funcionalidad del World Wide Web se ha ido restringiendo crecientemente por las limitaciones del HTML (y la complejidad del SGML).
- XML es un estándar emergente para Lenguajes de Marcas.
- Esta siendo desarrollado por un consorcio de casi 300 empresas bajo la guía del *World Wide Web Consortium* (W3C).
- 1^{ra} especificación: Febrero 1998.
- XML es visto por algunos como la herramienta que revolucionará al WWW (y de hecho lo está haciendo).

...XML

- Es un metalenguaje: “un conjunto de reglas que gobiernan el desarrollo de etiquetas o marcas únicas para codificar documentos”
- Es “extensible” (se puede usar para muchos propósitos)
- Menos complejo que SGML
- XML no tiene la intención de reemplazar a HTML, los dos son complementarios.
- Los dos estándares hacen énfasis en diferentes elementos de un documento:
 - *HTML maneja “formato y presentación de datos” de texto e imágenes*
 - *XML resuelve el problema de estructura y “significado o semántica de los datos”*
- XML introduce la posibilidad de manejo de información estructurada en el WWW
- Suministra **INFORMACIÓN**, no sólo páginas con texto e imágenes

...XML. Ejemplo sencillo

<poema>

```
<titulo> Lento pero viene </titulo>
<autor>
  <nombre> Mario </nombre>
  <apellido> Benedetti </apellido>
</autor>

<fecha_creación> 1986 </fecha_creación>

<formato> text/txt </formato>
```

Metadatos

<contenido>

Lento pero viene
el futuro real
el mismo que inventamos
nosotros y el azar

cada vez mas nosotros
y menos el azar...

...lento pero viene

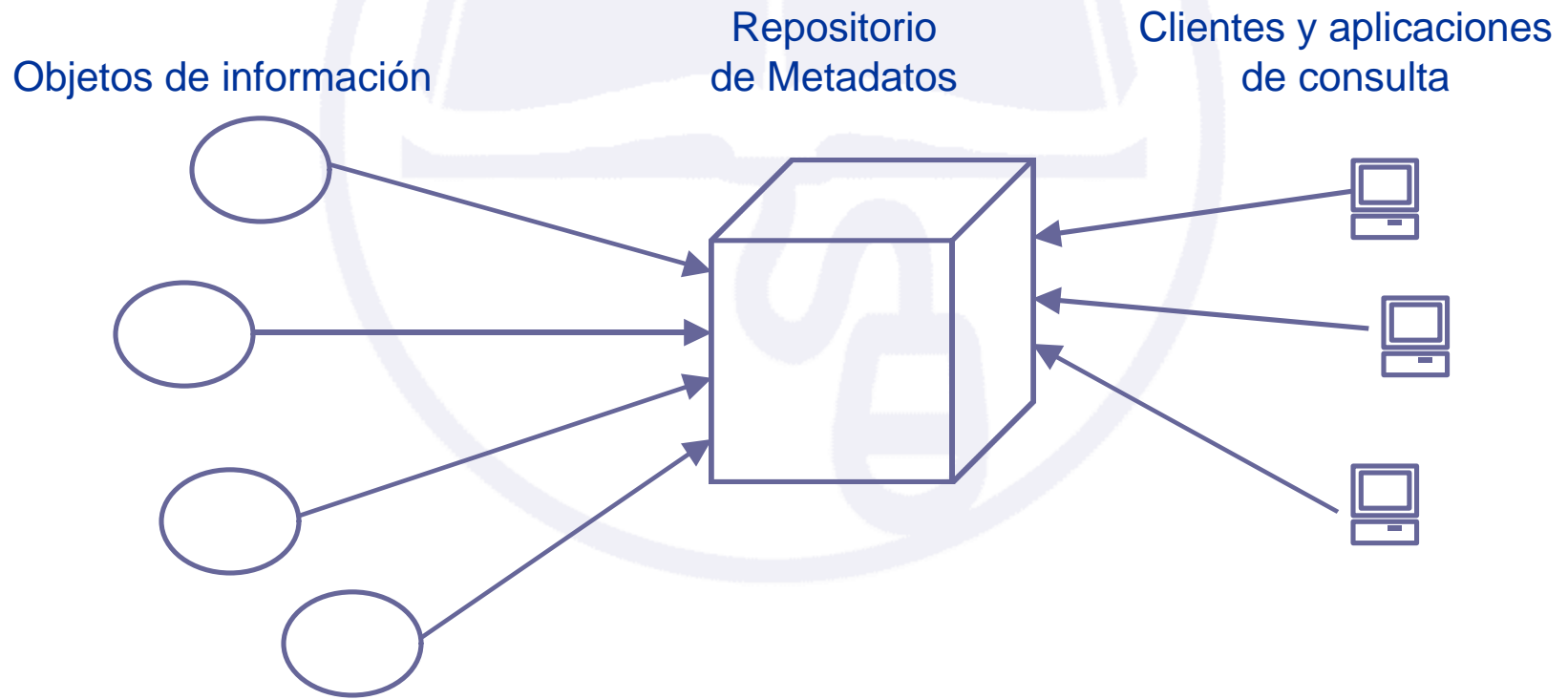
Datos

</contenido>

</poema>

Metadatos

Los **metadatos** son las instrucciones que describen el contenido, contexto, calidad, estructura y accesibilidad de un conjunto de datos u objetos de información.



XML y las Bibliotecas Digitales

- Muchas bibliotecas en Europa y Norteamérica están usando SGML
- Como XML ofrece más funcionalidad que HTML y menos complejidad que SGML, es natural que los bibliotecólogos consideren el uso de XML
- El primer campo donde las bibliotecas notarán y se beneficiarán de XML será en el manejo de recursos de información en el World Wide Web.
- Uso importante de XML dentro de las Bibliotecas:

Sistemas de Metadatos

...XML y las Bibliotecas Digitales

- Las colecciones electrónicas o digitales se volverán más y más importantes.
- A los bibliotecólogos se les presentará el reto de mantener actualizadas y organizadas estas crecientes colecciones de contenidos digitales.
- Allí entran en juego los estándares y lenguajes como XML

Compartimiento de Información

- Quedan muchos problemas por resolver en cuanto a estándares y lenguajes.
- Están apareciendo propuestas que organizarán el “desorden” actual.

Compartimiento de Información

Por ejemplo:

Estándares para Metadatos, contenidos estructurados y semántica de la información

- **Dublin Core, RDF**
- Estándares para objetos de información particulares (**ETD MS** por ejemplo)

...que comienzan a usar y aprovechar las facilidades e infraestructura de lenguajes como XML.

...Compartimiento de Información

También se hacen necesarios estándares de consulta, recolección (harvesting) e intercambio de información.

Ejemplos:

- **OAI: Open Archives Initiative**
- **OAI PMH: OAI Protocol for Metadata Harvesting**

...que son usados por muchos proyectos y organizaciones a lo largo del mundo:

- e-prints.org, OCLC (Online Computer Library Center, Inc.)
- NDLTD (Networked Digital Library of Theses and Dissertations)
- PubMed Central
- Archivo de e-prints de los Alamos...

Dublin Core y OAI

Dublín Core Metadata Initiative (DCMI):

La DCMI es una organización dedicada a promover la adopción de estándares de metadatos interoperables y a impulsar el desarrollo de vocabularios especializados de metadatos para describir recursos, que permitan la implementación de sistemas inteligentes de recuperación de datos.

...Dublin Core y OAI

...Dublín Core Metadata Initiative (DCMI):

A la DCMI se le ha encargado el continuo refinamiento de un “núcleo” (*core*) fundamental de tipos de propiedades y valores que proveen información semántica acerca de recursos en el Web, de la misma manera como un fichero de una biblioteca provee índices de las propiedades de un libro.

Los metadatos propuestos por el *Dublín Core* son usados para suplementar los métodos de búsqueda e indexación de metadatos basados en el Web existentes actualmente, sin importar si el recurso descrito es un documento electrónico o un objeto físico real.

...Dublin Core y OAI

Open Archives Initiative (OAI):

Desarrolla y promueve estándares de interoperabilidad que tienen el propósito de facilitar la diseminación eficiente de contenidos.

Tiene sus raíces en el esfuerzo para mejorar el acceso a archivos de e-prints y de difundir la producción académica.

Protocolos de consulta, intercambio y recolección definido por OAI son usados en muchos proyectos,

...Compartimiento de Información

Para qué todo esto? Si ya existen hace años SGML, o estándares como MARC, protocolos como Z39.50?

Se trata de que la información sea:

“entendible por computadoras”

(machine understandable)

...lo que provocara nuevos niveles de procesamiento, intercambio y recuperación automatizada de recursos de información entre instituciones (como las que ahora llamamos “bibliotecas”)

El Web Semántico

Todo lo anterior nos va llevando poco a poco a los...

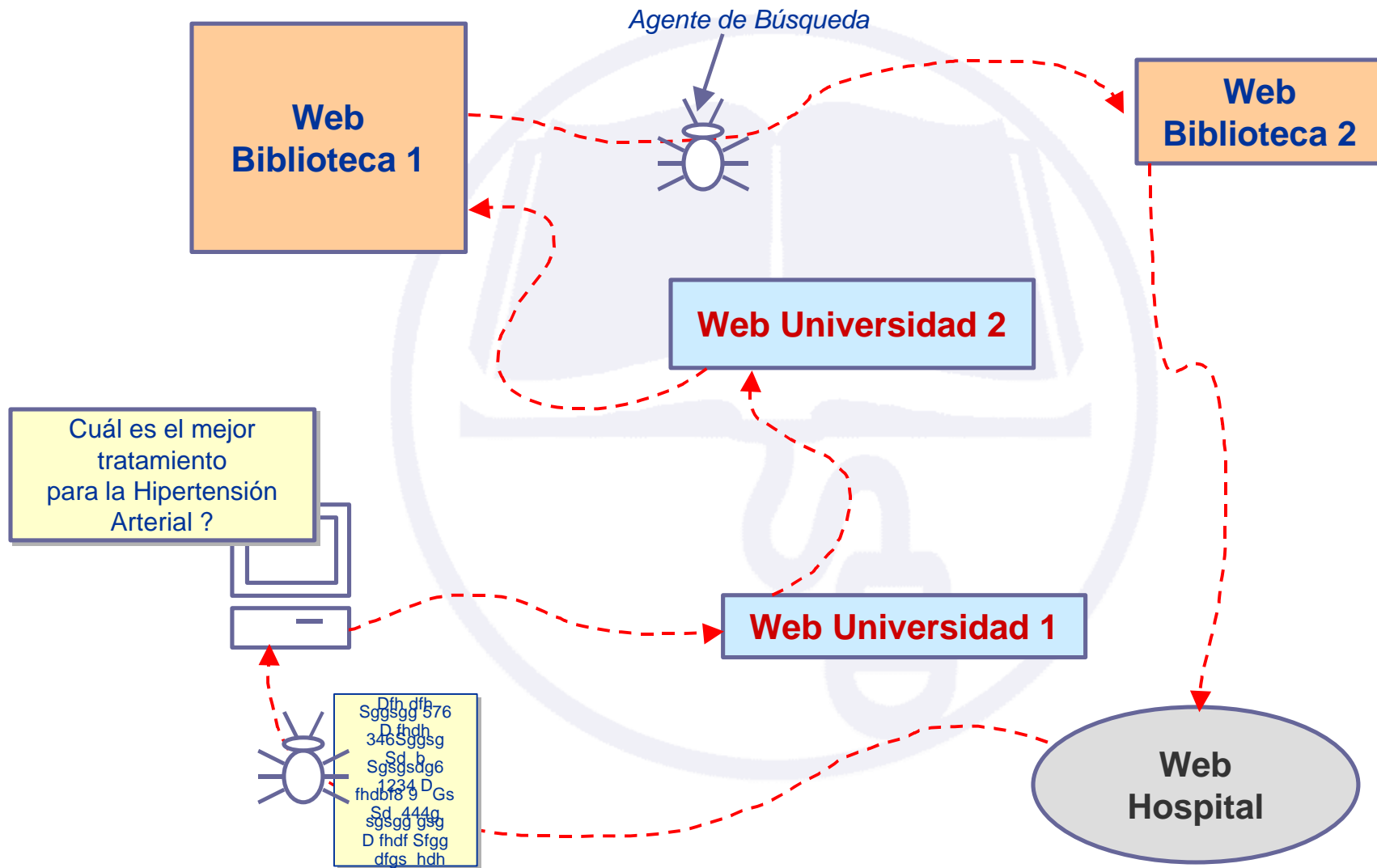
“Ambientes de Investigación Digitales”

..mencionados antes, y más ambiciosamente al:

“Web Semántico”

(Suponiendo que continúe la tendencia actual a usar el Web y tecnologías asociadas como puente a todos los recursos de información disponibles en la Red)

El Web Semántico



Herramientas “Futuras”

(Que ayudarán a construir las Bibliotecas del Futuro)

- Nuevos estándares (o uso más amplio de los propuestos actualmente)
 - Agentes de Software usarán principios de IA, redes neuronales, etc. para monitorear, administrar, gerenciar y asignar servicios y recursos de información.
 - Herramientas de comunicación permitirán el uso de Bibliotecas “ubicuas” (estarán presentes en todas partes y las usaremos sin que nos demos cuenta) ...
- ...(lo mismo que sucederá con el resto de Internet y los recursos que ofrece)

Para finalizar:

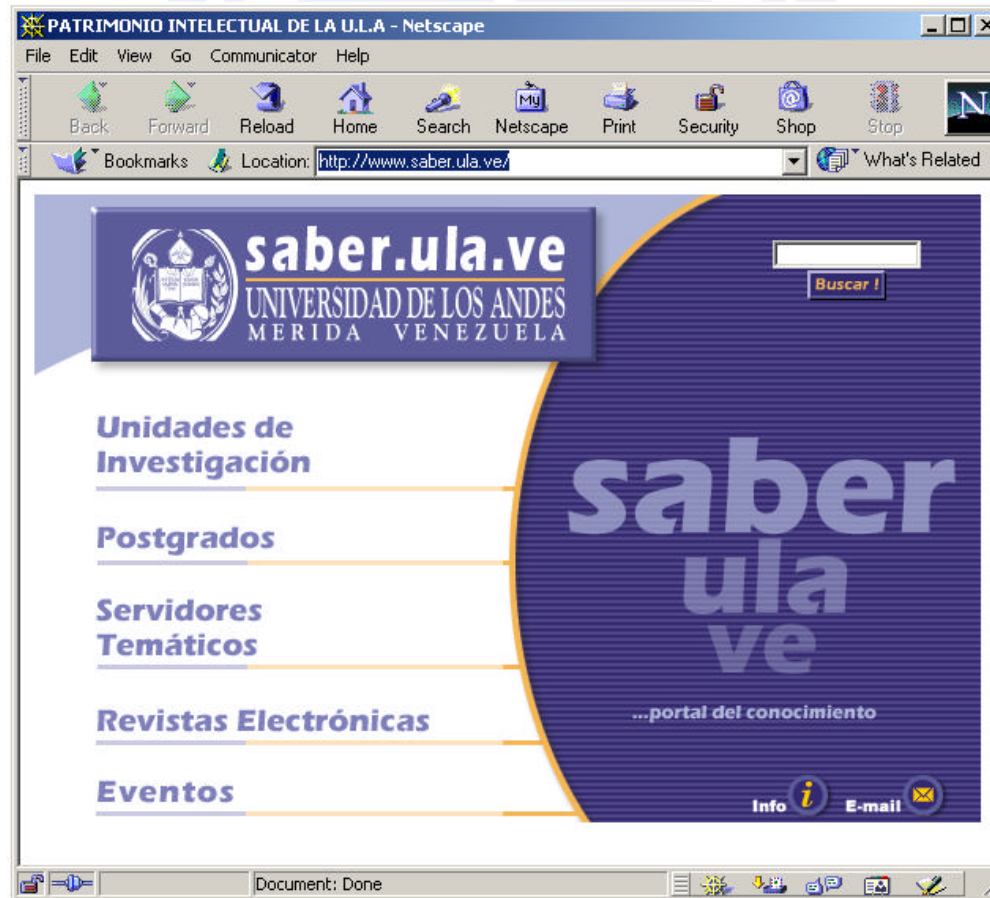
Muestra de nuestros esfuerzos locales y nacionales (e internacionales)

- Red de Datos de la Universidad de Los Andes (ULA)
- Taller Latinoamericano de Redes
- Talleres de TI y redes permanentes, programas de CISCO, Sun, etc.
- Proyecto Alejandría
- Proyecto SABER-ULA
- Proyecto Tesis Electrónicas-ULA
- Biblioteca Digital Andina (BDA)

Para finalizar:

...Muestra de nuestros esfuerzos locales y nacionales
(e internacionales)

- Proyecto SABER-ULA: www.saber.ula.ve



A manera de conclusión

Las verdaderas herramientas son los
estándares y protocolos
(maneras de comunicarse)

(Se oyen opiniones)...

Fuentes de información

- XML: www.xml.org, www.xml.com
- Dublin Core: www.purl.org, dublincore.org
- OAI: www.openarchives.org
- E-prints: www.eprints.org
- NDLTD, ETD: www.ndltd.org, etd.vt.edu
- Web Semántico:
www.semanticweb.org/
www.w3.org/2001/sw/
www.sciam.com/2001/0501issue/0501berners-lee.html
- Proyecto Alejandría: alejandria.ula.ve
- Proyecto SABER-ULA: www.saber.ula.ve

...Fuentes de información

- Edward Fox. *From Theory to Practice in Digital Libraries: 5S and Educational Applications (NDLTD, CSTC)*. Workshop on Digital Libraries. Albuquerque 07/1999
- Hussein Suleman, Edward Fox. *A Framework for Building Open Digital Libraries*. D-Lib Magazine. December 2001.
- Hussein Suleman, et al. *Networked Digital Library of Theses and Dissertations*. D-Lib Magazine. September 2001
- Luis Núñez. *Bibliotecas, autores, editores y estado*. Revista Investigación. Universidad de Los Andes.
URL: <http://www.saber.ula.ve/>
- Special Issue on Digital Libraries*. Comm. Of the ACM. May 2001. Vol. 44, Num. 5

Gracias por su atención !

Rodrigo Torrén Heeren
torrens@cecalc.ula.ve

Universidad de Los Andes
Unidad de Teleinformación
Edif. Gral. Masini. Av. 4, Piso 3, Of. A-32
E-mail: info@saber.ula.ve
Web: www.saber.ula.ve
Tel/Fax: +58 274 2524192